Structural, functional and genetic analyses of the actinobacterial transcription factor RbpA

Elizabeth A. Hubin[1]†, Aline Tabib-Salazar[2]†, Laurence Humphrey[2], Joshua E. Flack[1], Paul Dominic B. Olinares [3], Seth A. Darst[1], Elizabeth A. Campbell[1*], and Mark S. Paget[2*]

[1]Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY 10065, USA.

[2] School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK.

[3]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA.

Present address: A.T.-S., Medical Research Centre for Molecular Bacteriology and Infection, Imperial College London, London SW7 2AZ, UK; J.E.F., Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.

Classification: Biological Sciences/Biochemistry

Author contributions: E.A.H, S.A.D., E.A.C., and M.S.P. designed research; E.A.H., A.T.-S., L.H., J.E.F., P.D.B.O. performed research; E.A.C., P.D.B.O., S.A.D., E.A.C., and M.S.P. analyzed data; E.A.H. S.A.D., E.A.C., and M.S.P. wrote the paper.

The authors declare no conflict of interest.

Data deposition: Coordinates and structure factors for *Mycobacterium tuberculosis* RbpA/σ_2 have been deposited in the Protein Data Bank under the accession code 4X8K.

† Each author contributed equally to this manuscript

*Correspondence to: m.paget@sussex.ac.uk, elizabeth.campbell@rockefeller.edu

Abstract:

Gene expression is highly regulated at the step of transcription initiation, and transcription activators play a critical role in this process. RbpA, an actinobacterial transcription activator that is essential in *Mycobacterium tuberculosis* (*Mtb*), binds selectively to Group 1 and certain Group 2 σ factors. To delineate the molecular mechanism of RbpA, we show that the *Mtb* RbpA Sigma Interacting Domain (SID) and basic linker (BL) are sufficient for transcription activation. We also present the crystal structure of the *Mtb* RbpA-SID in complex with domain 2 of the housekeeping σ factor, σ^A. The structure explains the basis of σ selectivity by RbpA, showing that RbpA interacts with conserved regions of σ^A as well as the non-conserved region (NCR), which is present only in housekeeping σ factors. Thus, the structure is the first to show a protein interacting with the NCR of a σ factor. We confirm the basis of selectivity and the observed interactions using mutagenesis and functional studies. In addition, the structure allows for a model of the RbpA-SID in the context of a transcription initiation complex. Unexpectedly, the structural modeling suggests that RbpA contacts the promoter DNA, and we present *in vivo* and *in vitro* studies supporting this finding. Our combined data lead to a better understanding of the mechanism of RbpA function as a transcription activator.

Keywords: Actinobacteria/Mycobacteria/RbpA/Transcription/X-ray crystallography

Significance

Initiation of transcription in bacteria relies on a multisubunit RNA polymerase in concert with a dissociable σ subunit that confers promoter recognition and opening to reveal the DNA template strand. RbpA, a transcription activator unique to Actinobacteria and essential in *Mycobacterium tuberculosis*, associates tightly with sigma and is required for efficient initiation, although its mechanism of action is unclear. Here we solve the crystal structure of an *M. tuberculosis* σ/RbpA complex and present evidence indicating that RbpA activates transcription through novel contacts with promoter DNA. The work sheds light on the mechanism of transcription initiation by *M. tuberculosis* RNA polymerase, which is a proven antibiotic target.

\body

Introduction

Bacterial RNA polymerase (RNAP) comprises a catalytic core (subunit composition $\alpha_2\beta\beta'\omega$) that is active for transcription elongation but requires an additional dissociable subunit, the σ factor, for promoter-specific initiation (1, 2). All bacteria contain a single primary σ that is essential for viability and directs transcription of most genes during vegetative growth. Most bacteria also harbor alternative σ factors that can reprogram the RNAP to orchestrate adaptive responses to specific signals such as stress and morphological development (3). Primary σs can make up to four sequence specific contacts with promoter DNA via three conserved helical domains (σ_2, σ_3 and σ_4) that are spread over one face of the RNAP (4-8). Within each structural domain are defined regions of sequence similarity; i.e., the structural domain σ_2 comprises regions 1.2, 2.1, 2.2, 2.3 and 2.4 (9). The key interactions involve the σ_2 and σ_4 domains, which are spaced appropriately to contact the -10 and -35 promoter elements, respectively (6).

The vast majority of biochemical and genetic studies on bacterial transcription initiation have focused on *Escherichia coli* (*Eco*) RNAP and its primary σ, σ^{70}. Recently, it has emerged that the regulation of transcription initiation in the Actinobacteria phylum, which includes major pathogens, such as *Mycobacterium tuberculosis* (*Mtb*), and antibiotic producers, such as *Streptomyces* spp., is distinct from the *Eco* system by the dependence on two initiation factors, CarD and RbpA, neither of which is found in *Eco* (10-12).

In mycobacteria, the essential protein CarD has been shown to be present at most promoters *in vivo* and to function as a transcription activator *in vitro* (10, 13). More recently, *Mtb* CarD has been shown to activate transcription initiation by

stabilizing the RNAP open complex with promoters (RPo) by preventing collapse of the transcription bubble (14). CarD makes a direct protein/protein interaction with the RNAP β subunit β1-lobe, and structural models suggest it also contacts the upstream edge of the -10 promoter element DNA in RPo (13).

RbpA was originally discovered in *S. coelicolor* (*Sco*) where it is a major component of RNAP holoenzyme (11). RbpA is found in all Actinobacteria, and like CarD, is essential for growth in *Mtb* (11, 12). In comparison to CarD, much less is known about the RbpA structural mechanism.

The structural architecture of isolated RbpA has been defined by solution NMR (15, 16). A central core domain (RCD) comprises a β-barrel fold and is flanked by an unstructured 26 amino acid N-terminal tail and a C-terminal segment predicted to harbor two α-helices linked to the RCD by a 15 amino acid basic linker (BL; Fig. 1A). In the absence of core RNAP, RbpA can form a stable binary complex with the σ_2 domain of the primary σ factors of both *Sco* (σ^{HrdB}) and *Mtb* (σ^{A}) (15, 16). The RbpA/σ_2 interaction is mediated by the C-terminal segment (which we designate here the Sigma Interaction Domain, SID), and point mutations that disrupt σ binding also disrupt RbpA function (15). In addition to primary σ factors, RbpA interacts with certain Group 2 σ factors, σ^{B} in *Mtb* and σ^{HrdA} in *Sco*, but does not interact with Group 3 or Group 4 σ factors (15-17). RbpA is present at transcription initiation complexes *in vivo* and stimulates transcription *in vitro* from a wide range of *Sco* σ^{HrdB}-, *Sco* σ^{HrdA}-, *Mtb* σ^{A}-, and *Mtb* σ^{B}-dependent promoters (15, 18), but the mechanism for RbpA-mediated transcription activation is unknown.

Here, we show that the RbpA-BL and SID are sufficient for *in vitro* transcription activation by RbpA, and we determine the X-ray crystal structure of the *Mtb* RbpA/σ^{A}_2

complex, revealing the essential RbpA-SID/σ^{A}_{2} interactions, as well as representing the first structure of a protein interacting with the non-conserved insert found exclusively in housekeeping σ factors. From this result, we use a combination of structural modeling, mutagenesis, and *in vitro* and *in vivo* functional studies to probe the mechanistic basis of RbpA function. Our results suggest that in addition to the RbpA-SID/σ^{A}_{2} protein/protein interaction, RbpA/promoter DNA interactions are crucial in the role of RbpA as a transcription activator.

Results

The RbpA-BL and SID are sufficient for transcription activation

While the essential role of the RbpA-SID in RbpA function is clear, the roles of the other RbpA structural elements (N-terminal tail, RCD, BL) are not. We therefore tested truncated derivatives of *Mtb* RbpA, N-terminally fused to SUMO to improve stability, for transcription activation function using a mycobacterial transcription system and an *Mtb vapB10* (Rv1398c) promoter template. The *vapB10p* promoter (19) is used here because it exhibits strong dependence on RbpA *in vitro*. As expected, an *Mtb* RbpA deletion mutant (containing only residues 1–71) lacking the BL and SID was unable to activate transcription *in vitro* (Fig. 1B, lanes 8-9). By contrast, the RbpA-SID and much of the BL (residues 72–C-terminus) showed partial *in vitro* function compared to full-length RbpA with an equivalent N-terminal SUMO fusion (Fig. 1B, lanes 10-11.). We conclude that the RbpA-BL and SID are necessary and sufficient for at least partial RbpA *in vitro* transcription activation, but that other elements of RbpA (N-terminal tail and/or RCD) are required for full activity.

X-ray crystal structure of the *Mtb* RbpA/σ^{A}_{2} complex

To provide insight into RbpA function as a transcription activator, we determined the crystal structure of the *Mtb* RbpA/σ^{A}_{2} complex to 2.2 Å resolution (Figs. 1C, S1A, Table S2). Although crystallization trials were set up with a purified complex of a His_6-SUMO-σ^{A}_{2} fusion (containing σ^{A}_{2} residues 224-364) and full-length RbpA (1-111), MALDI-TOF analysis of the crystal contents revealed that both proteins were N-terminally proteolytically degraded, resulting in crystals containing multiple σ^{A}_{2} fragments (lacking up to 18 N-terminal residues; see supplemental methods) and multiple RbpA

fragments (all including the BL and SID; see supplemental figures) (Fig. 1A). Electron density maps revealed the RbpA-SID (77-108) bound to σ^{A}_{2} (242-363) (Fig. 1C). Although the MALDI-TOF analysis indicated that the mixture of RbpA fragments crystallized included intact RbpA molecules, additional electron density was absent, and the rest of RbpA was presumed disordered or absent.

The structure shows that the RbpA-SID comprises two α-helices (α1, α2; Fig. 1C), confirming previous sequence-based structural predictions (15-17). Four residues of the 15-residue RbpA-BL connecting the RbpA-SID to the RCD are also visible in the structure. Both α-helices of the RbpA-SID contact σ^{A}_{2}, forming an intermolecular interface with a buried surface area of 948 $Å^2$.

The RbpA-SID makes extensive contacts with residues from conserved regions of σ^{A}_{2} (1.2 and 2.3) as well as the NCR (Figs. 1C, 2A-C, S1). Interacting with the primary σ-NCR is a property RbpA shares with the unrelated *Chlamydia trachomatis* (*Ctr*) transcription factor GrgA, which binds to the *Ctr* σ^{66}-NCR (20). Moreover, although structurally distinct from RbpA, the holoenzyme assembly factor Crl from enteric bacteria interacts with the equivalent region of the Group 2 sigma factor σ^{S} (21, 22). The co-crystal structure of RbpA/σ^{A}_{2} represents the first structure of an activator (or any protein) interacting with the NCR of a housekeeping σ.

Previous studies identified two conserved arginine residues (*Sco* RbpA R89 and R90, corresponding to *Mtb* R88 and R89*;* Fig. 2B) that are critical for σ binding (15). The structure shows that these two residues form extensive electrostatic interactions with σ^{A} (Fig. 2C), explaining the mutagenesis results. *Mtb* RbpA-R88, in α1 of the SID, forms a salt bridge with σ^{A} residue E254, while RbpA-R89, located in the linker between the two SID α-helices, makes salt bridges with both E254 and D336 of σ^{A}.

A cluster of conserved hydrophobic RbpA residues (L94, L97, and L98; Figs. 2B, C), located on RbpA-SID α2, make extensive van der Waals contacts with residues of σ^A (Figs. 2C, S1). To determine if the interactions observed in the crystal structure are of wider importance, we performed bacterial two-hybrid (BTH) assays with a *Sco* RbpA mutant in which the residues corresponding to *Mtb* RbpA L97 and L98 (*Sco* RbpA V98 and L99; Fig. 2B) were changed to alanine. The results reveal that these branched hydrophobic residues are necessary for *Sco* RbpA/σ^{HrdB} binding (Fig. S2A,B).

Identification of RbpA residues involved in σ selectivity

In addition to binding primary σ factors, RbpA binds some Group 2 σ factors (*Mtb* σ^B and *Sco* σ^{HrdA}) but not others (*Sco* σ^{HrdC} and σ^{HrdD}), and does not interact with the more diverse Group 3 (e.g. *Mtb* σ^F) or Group 4 (e.g. *Sco* σ^R) σ factors (15, 17). The RbpA-SID/$\sigma^A{}_2$ structure provides a basis to understand the σ selectivity of RbpA. An alignment of σ_2 amino acid sequences of *Mtb* σ^A and σ^B with *Mtb* σ^F and *Sco* σ^{HrdC} (Fig. 2A) revealed that out of 21 *Mtb* σ^A residues contacting RbpA, 16 were identical in *Mtb* σ^B while only four were identical in *Mtb* σ^F, explaining why RbpA is specific to Group 1 and 2 σ factors (17). Although there was extensive conservation between *Mtb* σ^A and *Sco* σ^{HrdC} among most RbpA-interacting residues, there were several positions in *Sco* σ^{HrdC} that were substituted with physico-chemically dissimilar amino acids that might impede RbpA binding (e.g. σ^A E248/σ^{HrdC} R59, σ^A K251/σ^{HrdC} T62, σ^A Y258/σ^{HrdC} R69). We tested this idea by mutating *Mtb* σ^A residues E248 and K251 in σR1.2 and L257and Y258 in the NCR, changing each position to the equivalent residue in *Sco* σ^{HrdC} (Fig. 2A). $\sigma^{A(RT)}$ contained the mutations E248R and K251T, $\sigma^{A(VR)}$ contained L257V and Y258R mutations, and $\sigma^{A(RTVR)}$ had all four positions altered to the equivalent positions in *Sco*

σ^{HrdC}. BTH assays suggested that all three mutants were defective in RbpA binding (Fig. 2D). To support this, $\sigma^{A(VR)}$ and $\sigma^{A(RTVR)}$ mutants were overexpressed, purified, and tested by *in vitro* transcription on the *vapB10p* promoter template. In the absence of RbpA, $\sigma^{A(VR)}$ and $\sigma^{A(RTVR)}$ directed basal transcription activity to the same extent as wild type σ^{A}. However, neither was responsive to RbpA, confirming that the interaction between σ and RbpA is crucial for the transcription activation role of RbpA (Fig. 2E).

RbpA interacts directly with promoter DNA

To gain insight into the role of RbpA in transcription initiation by RNAP holoenzyme, we generated a structural model of the complex between RbpA and RPo by superimposing the conserved regions of σ^{A}_2 from the *Mtb* RbpA-SID/σ^{A}_2 complex (Fig. 1C) onto the corresponding regions of *Taq* σ^{A}_2 in an RPo model (6, 23) (0.681 Å root-mean-square deviation over 93 Cα atoms) (Fig. 3), resulting in an RbpA/RPo model with no steric clashes. In the model, the RbpA-SID/σ^{A}_2 interaction positions the RbpA-BL (Fig. 1A) on the minor-groove side of the duplex promoter DNA just upstream of the -10 element, with absolutely conserved RbpA-R79 (Fig. 2B) positioned to interact with the non-template strand (nt-strand) DNA phosphate backbone at the -14 position with respect to the transcription start site at +1 (Fig. 3B). In addition, absolutely conserved RbpA-M84 (Fig. 2B) in α1 is positioned to play a role in DNA binding through van der Waals interactions (Fig. 3B). Also note that the next three residues N-terminal to the modeled portion of RbpA are all lysines (K73, K74, K76), which may also play a role in forming electrostatic interactions with the DNA phosphate backbone (Fig. 3B).

To test the hypothesis of an RbpA/DNA interaction, we used formaldehyde cross-linking, which links atoms 2 Å apart (24). In order to form stable RNAP/DNA

complexes, we used a fork-junction template (6, 25) comprising the *vapB10* promoter (VFJ-28; Figs. 4A, S3A) along with an anti-consensus control (VFJ-28anti; Fig. S3A). The inclusion of RbpA in cross-linking reactions led to the appearance of a new band (Fig. 4A, lane 3), which was confirmed to be an RbpA-DNA crosslink since a slower migrating band appeared when the assay was repeated with SUMO-RbpA (Fig. 4A, lane 4). Furthermore, RbpA-DNA cross-linking was detected with SUMO-RbpA(72-111), which includes the RbpA-BL and SID, but not with SUMO-RbpA(1-71) lacking the RbpA-SID essential for σ^A binding and activity (Fig. 4A, lanes 5-6). This finding is consistent with the ability of SUMO-RbpA(72-111), but not SUMO-RbpA(1-71), to activate transcription (Fig. 1B). No crosslinking was detected with VFJ-28anti or with reactions that lacked core RNAP (Fig. S3B), indicating that RbpA/$\sigma^A{}_2$ interactions with DNA occur only in the context of RNAP holoenzyme/promoter complexes.

An RbpA/promoter DNA interaction would be predicted to increase the overall affinity of RNAP for promoter DNA. To test this hypothesis, we measured RNAP binding to a Cy3-labeled duplex *vapB10* promoter template (dsVapB; Fig. S4A) using a fluorescence anisotropy assay (26). We found that addition of RbpA decreased the dissociation constant, K_D, for binding to the promoter DNA nearly two-fold (Fig. 4B), consistent with the modest activation activity of RbpA in abortive initiation (Fig. S4B) and run-off (Fig. 1B) transcription assays. On the other hand, the RbpA-R79A mutant had no significant effect on RNAP promoter binding (Fig. 4B), consistent with the hypothesis that RbpA-R79 plays an important role in DNA binding (Fig. 3B). Wild type RbpA had no significant effect on RNAP binding to Cy3-labeled single-stranded DNA comprising only the -10 and discriminator elements (Table S3), supporting the structure-based hypothesis that the effect of RbpA on RNAP-promoter binding is

through the interactions of RbpA with duplex DNA just upstream of the -10 element (Fig. 3).

DNA interaction via RbpA-R79 is critical for transcription activation

We hypothesized that an interaction between RbpA and DNA upstream of the -10 element might underlie the transcription activation function of RbpA. To test this we used multi-round *in vitro* transcription assays and found that while *Mtb* RbpA-M84A retained partial activity, *Mtb* RbpA-R79A completely failed to stimulate transcription from a *vapB10* promoter template (Fig 4C, lanes 5-6). In order to investigate the importance of these residues *in vivo*, we performed experiments in *Sco* where RbpA is required for normal growth but not essential for viability. The *Sco* mutants *rbpA-R80A* and *rbpA-M85A* (corresponding to *Mtb* RbpA R79A and M84A) were cloned into an integrative vector and used to transform the Δ*rbpA Sco* mutant S129 (15). The *rbpA-R80A* and *rbpA-M85A* alleles retarded growth leading to smaller colonies on agar plates compared to the control *rbpA-WT* strain (Fig. S5). These data indicate that the conserved *Mtb*/*Sco* RbpA R79/80 and M84/85 residues are critical for normal *rbpA* function *in vivo*.

Discussion

We propose that a key role of the RbpA-SID/$\sigma^{A}{}_{2}$ interaction in transcription activation is to position the RbpA-BL near the upstream edge of the -10 promoter element to facilitate its interaction with the DNA phosphate backbone. Adding favorable protein/DNA contacts to the transcription initiation complex would potentially stabilize a transcription initiation intermediate, which is consistent with findings that RbpA

activates transcription primarily by stimulating RPo formation (15, 17). The role of RbpA-R79, which structural modeling predicts could contact the nt-strand DNA phosphate backbone at the -14 position (Fig. 3B), appears to be particularly significant. This residue is absolutely conserved among RbpA orthologues (Fig. 2B), stabilizes the binding of RNAP to promoter DNA (Fig. 4B), and is required for RbpA transcription activation *in vitro* (Fig. 4C) and for full RbpA function *in vivo* (Fig. S5). Strikingly, the RbpA-BL also contains three lysine residues (K73, K74, K76; Figs. 2B, 3B), which, although not modeled in the structure, might also contribute to interactions with the DNA phosphate backbone around the -14/-15 positions (Fig. 3B).

Although the RbpA-BL and SID together are sufficient for partial transcription activation *in vitro*, full activation requires full-length RbpA (the N-terminal tail/RCD) (Fig. 1B). We suggest that the N-terminal tail and/or RCD likely form additional interactions with RNAP. Interactions between RbpA and core RNAP have been proposed previously, and putative interactions have been mapped to two widely spaced regions of RNAP (18, 27). However, our RbpA-SID/RPo structural model is completely incompatible with previous claims that RbpA interacts with RNAP either in the active-site channel near the Rif binding site (27, 28) (Fig. S6A) or with a different region of the β-subunit (18) (Fig. S6B). Because of the location and orientation of the N-terminus of the RbpA-SID in our structural model, it would be impossible for any part of RbpA to span the distance from the N-terminus of our model to either of the putative sites (Fig. S6). The binding site near the Rif pocket was originally inferred based on functional data (28), then refined on the basis of a single crosslink (27). The other β-subunit determinant was identified using cleavage experiments through hydroxyl-radicals generated from Fe-BABE attached to the lone Cys-residue of the RbpA-RCD. In any case, whether or not the RbpA-RCD or N-terminal tail binds RNAP at all has not been

established, and if they do, our RbpA-SID/RPo model suggests that the likely RNAP binding determinant is on the β' subunit on the top of the clamp domain (Fig. S6A). More experiments will be required to resolve these inconsistencies. A possible interaction with the RNAP clamp domain has also been proposed for Crl, a σ^S-RNAP holoenzyme assembly factor found in *Eco* and other γ-proteobacteria. Crl is thought to bind to a surface-exposed region of $\sigma^S{}_2$ that is equivalent to the surface of $\sigma^A{}_2$ bound by RbpA (21).

Both RbpA and CarD are essential transcription activators in *Mtb* (10, 12). CarD is associated with essentially all σ^A promoters in *M. smegmatis in vivo* (13), and results suggest that RbpA also likely functions at all σ^A promoters *in vivo* (11). Therefore, it seems likely that RbpA and CarD may function simultaneously on the same RNAP/promoter complexes. Indeed, structural modeling indicates that RbpA and CarD interactions with RPo are completely compatible, each interacting with DNA upstream of the -10 element but from opposite sides of the DNA. Future studies will address the influence of these two transcription regulators on each other and their possible role in modulating global transcription patterns in the Actinobacteria.

Materials and Methods:

Full details of the methods used are presented in the **SI Materials and Methods**

Strains, plasmids and growth conditions. Strains, plasmids and oligonucleotides are described in Table S1. *Sco* A3(2) strains were conjugated with *Eco* ET12567 (pUZ8002) and cultivated on mannitol-soya (MS) agar or in Yeast-Extract Malt-Extract (YEME) liquid medium (29). Bacterial two hybrid analysis was carried out using *Eco* BTH101 derivatives as described (15).

Protein Expression and Purification. For crystallography, σ_2 (codons 224-364) and RbpA (codons 1-111) from *Mtb* were cloned into a His_6 pET SUMO co-expression vector, and proteins were over-expressed in *Eco* BL21(DE3) cells (Novagen). The complex was purified by Ni-affinity chromatography and size exclusion chromatography. For fluorescence anisotropy, *Mtb* RbpA and *Mtb* RbpA-R79A were cloned into a His_6 pET28-based vector, and were expressed and purified as described for the complex. Core RNAP, σ^A, and CarD were over-expressed and purified as described (14). For *in vitro* transcription and cross-linking assays, *Mtb* RbpA and SUMO-RbpA fusions and mutant derivatives were cloned into pET20b (Novagen) and pET SUMO (Life Technologies), respectively, overexpressed in *Eco* BL21(pLysS), and purified. The σ^A and mutant derivatives were cloned into pET15b and purified from *Eco* BL21(pLysS).

X-ray Structure Determination of the RbpA/$\sigma^A{}_2$ Complex. The purified complex was concentrated to ~15 mg/mL by centrifugal filtration, and crystals were grown at 22 °C by sitting drop vapor diffusion. MALDI-TOF MS was performed on the washed, dissolved crystals to determine the protein content using a Spiral TOF JMS-S3000 (JEOL,

Tokyo, Japan). The structure was solved using an *Mtb* σ^{A}_{2} homology model based on a structure of *Taq* σ^{A}_{2} with the non-homologous NCR removed [1KU2;(4)] (Table S2).

Fluorescence Anisotropy. Fluorescence measurements were performed using an Infinite M1000pro plate reader (Tecan). Reactions were performed with Cy3-labeled duplex *vapB10* DNA (Fig. S4A, Tri-link for labeled nt-strand, Oligos Etc. for unlabeled template strand) extending from -36 to +1, or single stranded Cy3 labeled DNA (Table S3, Tri-link) extending from -12 to +1. RNAP binding to duplex DNA gave poor signal-to-noise data without the addition of CarD. Recent studies have shown that *Mtb* RNAP makes relatively unstable open complexes and displays short half lives on duplex promoter DNA without CarD (14). Therefore, these assays were performed in the presence of CarD. Data were analyzed using Prism software.

Formaldehyde cross-linking. The crosslinking assay was modified from (24) and used as substrate *vapB10* fork-junction DNA (VFJ-28) including an 18 bp double stranded region extending from -28 to -12 followed by a single-stranded extension from -11 to -3 on the non-template strand (Fig. S3A). Cross-linking was performed for 2 min in the presence of 60 mM formaldehyde.

Acknowledgments

We thank B. Bae for aiding in the harvesting and freezing of crystals. We thank R.L. Landick and R. Mooney for sharing plasmids for overexpressing mycobacterial RNAP and σ^A. The use of the Formulator and Phoenix robots in the Rockefeller University Structural Biology Resource Center was made possible by D. Oren and Grant Number 1S10RR027037-01 from the National Center for Research Resources of the NIH. This work was supported by BBSRC grant BB/I003045/1, NIH grants U54 GM103511, P41 GM109824 and P41 GM103314 (P.D.B.O.) and R01 GM053759 (S.A.D.), and R01 GM114450 (E.A.C.).

References

1. Burgess RR, Travers AA, Dunn JJ, Bautz EK (1969) Factor stimulating transcription by RNA polymerase. *Nature* 221(5175):43–46.

2. Murakami KS, Darst SA (2003) Bacterial RNA polymerases: the wholo story. *Current Opinion in Structural Biology* 13(1):31–39.

3. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* 57:441–466.

4. Campbell EA, et al. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Molecular Cell* 9(3):527–539.

5. Murakami KS, Masuda S, Darst SA (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science* 296(5571):1280–1284.

6. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296(5571):1285–1290.

7. Vassylyev DG, et al. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417(6890):712–719.

8. Zhang Y, et al. (2012) Structural basis of transcription initiation. *Science* 338(6110):1076–1080.

9. Lonetto M, Gribskov M, Gross CA (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology* 174(12):3843–3849.

10. Stallings CL, et al. (2009) CarD Is an Essential Regulator of rRNA Transcription Required for Mycobacterium tuberculosis Persistence. *Cell* 138(1):146–159.

11. Newell KV, Thomas DP, Brekasis D, Paget MSB (2006) The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on Streptomyces coelicolor. *Molecular Microbiology* 60(3):687–696.

12. Forti F, Mauri V, Dehò G, Ghisotti D (2011) Isolation of conditional expression mutants in Mycobacterium tuberculosis by transposon mutagenesis. *Tuberculosis (Edinb)* 91(6):569–578.

13. Srivastava DB, et al. (2013) Structure and function of CarD, an essential mycobacterial transcription factor. *Proceedings of the National Academy of Sciences* 110(31):12619–12624.

14. Davis E, Chen J, Leon K, Darst SA, Campbell EA (2015) Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucl Acids Res* 43(1):433–445.

15. Tabib-Salazar A, et al. (2013) The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. *Nucl Acids Res* 41(11):5679–5691.

16. Bortoluzzi A, et al. (2013) Mycobacterium tuberculosis RNA polymerase binding protein A (RbpA) and its interactions with sigma factors. *J Biol Chem*.

17. Hu Y, Morichaud Z, Sudalaiyadum Perumal A, Roquet-Baneres F, Brodolin K (2014) Mycobacterium RbpA cooperates with the stress-response σB subunit of RNA polymerase in promoter DNA unwinding. *Nucl Acids Res* 42(16):10399–10408.

18. Hu Y, Morichaud Z, Chen S, Leonetti JP, Brodolin K (2012) Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the A-containing RNA polymerase holoenzyme. *Nucl Acids Res* 40(14):6547–6557.

19. Cortes T, et al. (2013) Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. *Cell Rep* 5(4):1121–1131.

20. Bao X, Nickels BE, Fan H (2012) Chlamydia trachomatis protein GrgA activates transcription by contacting the nonconserved region of σ66. *Proceedings of the National Academy of Sciences* 109(42):16870–16875.

21. Banta AB, et al. (2013) Key features of σS required for specific recognition by Crl, a transcription factor promoting assembly of RNA polymerase holoenzyme. *Proceedings of the National Academy of Sciences* 110(40):15955–15960.

22. Banta AB, et al. (2014) Structure of the RNA polymerase assembly factor crl and identification of its interaction surface with sigma s. *Journal of Bacteriology* 196(18):3279–3288.

23. Feklistov A, Darst SA (2011) Structural Basis for Promoter -10 Element Recognition by the Bacterial RNA Polymerase σ Subunit. *Cell* 147(6):1257–1269.

24. Brodolin K, Mustaev A, Severinov K, Nikiforov V (2000) Identification of RNA polymerase beta' subunit segment contacting the melted region of the lacUV5 promoter. *J Biol Chem* 275(5):3661–3666.

25. Guo Y, Gralla JD (1998) Promoter opening via a DNA fork junction binding activity. *Proceedings of the National Academy of Sciences* 95(20):11655–11660.

26. Quantification of protein-protein interactions using fluorescence polarization. (1999) Quantification of protein-protein interactions using fluorescence polarization. 19(2):222–233.

27. Dey A, Verma AK, Chatterji D (2011) Molecular insights into the mechanism of phenotypic tolerance to rifampicin conferred on mycobacterial RNA polymerase by MsRbpA. *Microbiology* 157(Pt 7):2056–2071.

28. Dey A, Verma AK, Chatterji D (2010) Role of an RNA polymerase interacting protein, MsRbpA, from Mycobacterium smegmatis in phenotypic tolerance to

rifampicin. *Microbiology* 156(Pt 3):873–883.

29. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood D (2000) Practical Streptomyces Genetics (The John Innes Foundation, Norwich, United Kingdom).

30. Cadene M, Chait BT (2000) A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Analytical chemistry*.

31. MALDI sample preparation: the ultra thin layer method. (2007) MALDI sample preparation: the ultra thin layer method. (3):192.

32. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 267:307–326.

33. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.

34. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.

35. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 372(3):774–797.

36. Paget MS, Chamberlin L, Atrih A, Foster SJ, Buttner MJ (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in Streptomyces coelicolor A3(2). *Journal of Bacteriology* 181(1):204–211.

37. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct

selective high-level expression of cloned genes. *Journal of Molecular Biology* 189(1):113–130.

38. Karimova G, Ullmann A, Ladant D (2001) Protein-protein interaction between Bacillus stearothermophilus tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *J Mol Microbiol Biotechnol* 3(1):73–82.

39. Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* 4(4):269–275.

40. Karplus PA, Diederichs K (2012) Linking Crystallographic Model and Data Quality. *Science* 336(6084):1030–1033.

41. O'Connor TJ, Kanellis P, Nodwell JR (2002) The ramC gene is required for morphogenesis in Streptomyces coelicolor and expressed in a cell type-specific manner under the direct control of RamR. *Mol Microbiol* 45(1):45-57.

Figure Legends

Fig. 1. Structural and functional analyses of RbpA-SID /σ^{A}_{2}. (A) Schematic of RbpA and σ^{A} domains. Domain 2 of *Mtb* σ^{A} is shown in orange with the NCR in cranberry, and the remaining regions colored gray. The SID and basic linker of RbpA are colored purple, and the remaining regions colored gray. Regions visible in the crystal structure are flanked by dashed lines. (B) Multiple round *in vitro* run-off transcription reactions using the *vapB10* promoter template. Reactions contained core *Mbo* RNAP (50 nM), σ^{A} (250 nM), and RbpA derivatives (500 nM, or 1.25 μM), as indicated. Lanes correspond to full length RbpA (residues 1-111, lanes 4 and 5), SUMO-RpbA full length (RbpA residues 1-111 fused to SUMO, lanes 6 and 7), SUMO-RbpA-RCD (RbpA residues 1-71 fused to SUMO, lanes 8 and 9) and SUMO-RbpA-SID (RbpA residues 71-111 fused to SUMO, lanes 10 and 11). A graphical representation based on duplicate data sets (standard deviation indicated) is illustrated above, normalized to the data obtained with σ^{A} in the absence of RbpA. (C) Crystal structure of *Mtb* RbpA-SID in complex with σ^{A}_{2} is shown in ribbon and colored as in (A).

Fig. 2. The X-ray crystal structure of the *Mtb* RbpA/σ^{A}_{2} complex explains the σ specificity of RbpA. (A) Amino acid sequence alignment of conserved regions 1.2 to 2.3 in σ_2 domains of *Mtb* σ^{A}, σ^{B}, and σ^{F}, and *Sco* σ^{HrdC}. The following amino acids, written in single letter nomenclature, were considered homologous and grouped as indicated within parenthesis: (VLMA), (RK), (ED), (YFW), (TS) and (QN). Remaining residues (H, C, P and G) were not grouped. The σ^{A} residues that contact RbpA are indicated by red

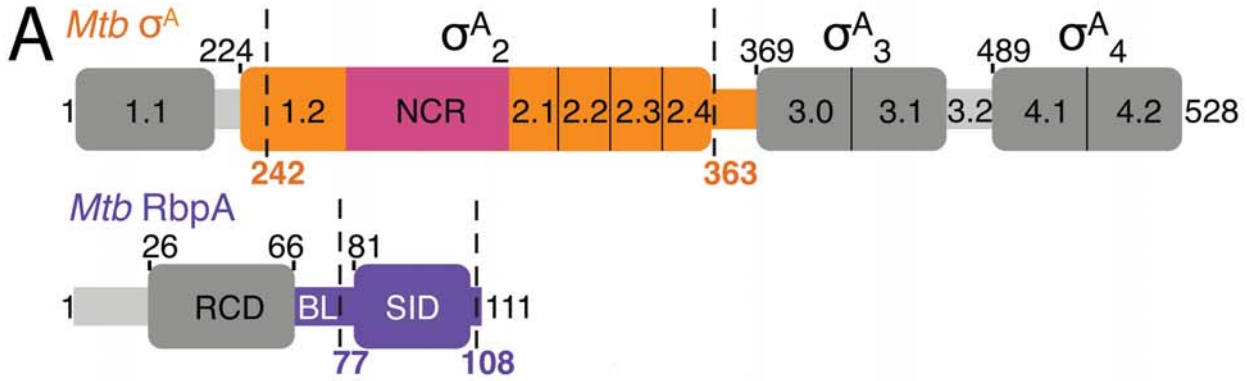
dots, with those that are targeted for mutagenesis outlined in black. (B) Alignment of RbpA BL and SID from diverse representatives of the order Actinobacteria. Mtb, *M. tuberculosis*; Sco, *S. coelicolor* A3(2), Fsp, *Frankia sp*.; Krd, *Kineococcus radiotolerans*; Cac, *Catenulispora acidiphila*; Mau, *Micromonospora aurantiaca*; Jga, *Jiangella gansuensis*; Gar, *Glycomyces arizonensis*; Agr, *Actinomyces graevenitzii*. Residues were considered in the same conservation group as (A). Green dots indicate residues that are suggested by the RbpA/RP_o model to come in close contact with DNA; blue dots highlight conserved arginine residues demonstrated important for binding σ (Tabib-Salazar et al., 2013); white dots indicate residues forming a hydrophobic interface between RbpA and σ and shown important for binding σ (this study). (C) Highlighted RbpA-SID residues required for σ^A binding. RbpA residues R88 and R89 form a cluster of salt bridges with σ^A residues E254 and D336 (selected interactions indicated by dotted lines). RbpA L94, L97, and L98 make van der Waals interactions with σ^A residues Y258, A259, I284 and K251. (D) BTH analysis of *Mtb* RbpA interactions with *Mtb* σ^A mutants predicted to be defective in RbpA binding. Selected σ^A residues that contact RbpA were changed to equivalent residues in *Sco* σ^{HrdC} as described in the text. The coding sequences of *rbpA* were fused to the T18 subunit of *Bordetella pertussis* adenylate cyclase in pUT18, whereas the sigma factor genes were fused to the T25 subunit in pKT25 (15). β-galactosidase assays were performed in triplicate (standard deviations indicated), and results are presented as % Miller units relative to wild-type *Mtb* σ^A. (E) RbpA activation of the *vapB10* promoter using σ^A mutants highlighted in (A). RbpA activation on mutant sigmas $\sigma^{A(VR)}$ or $\sigma^{A(RTVR)}$ was compared to wild-type $\sigma^{A(WT)}$ using multiple round *in vitro* run-off transcription reactions. Reactions contained core *Mbo* RNAP (50 nM), wild-type $\sigma^{A(WT)}$, $\sigma^{A(VR)}$, or $\sigma^{A(RTVR)}$ (each 250 nM) and the presence or absence of RbpA (500 nM), as indicated. A graphical representation based on

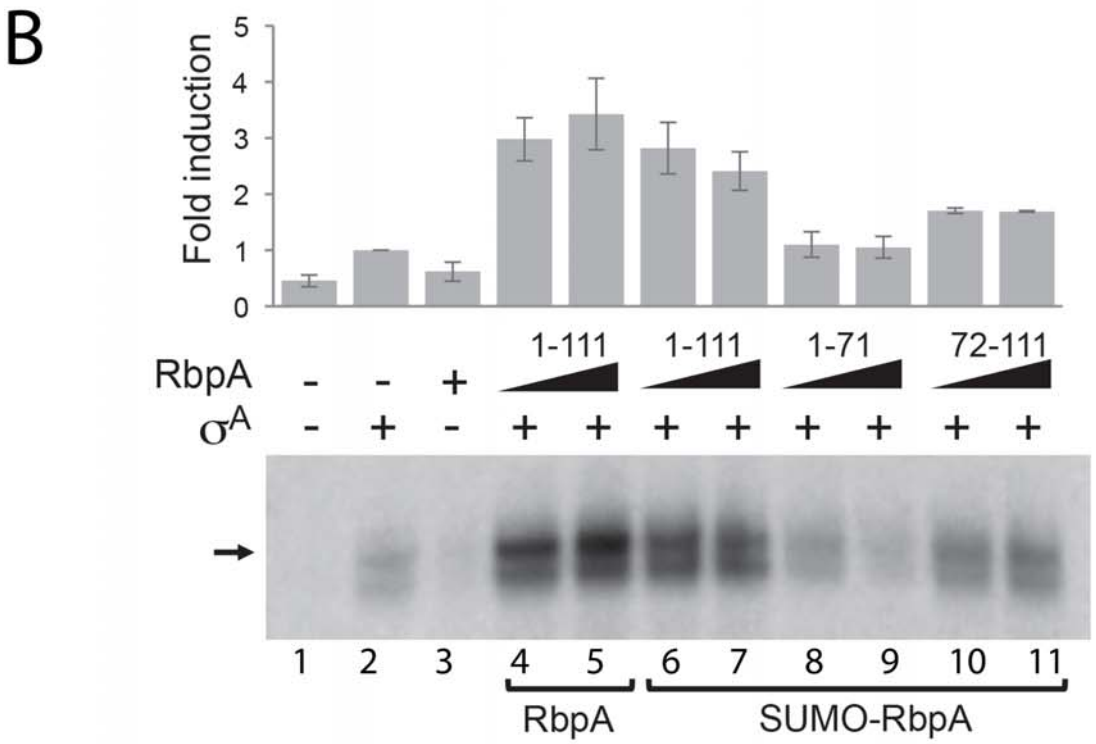
triplicate data sets, with standard deviation indicated, is illustrated above, normalized to the data obtained with $\sigma^{A(WT)}$ in the absence of RbpA.

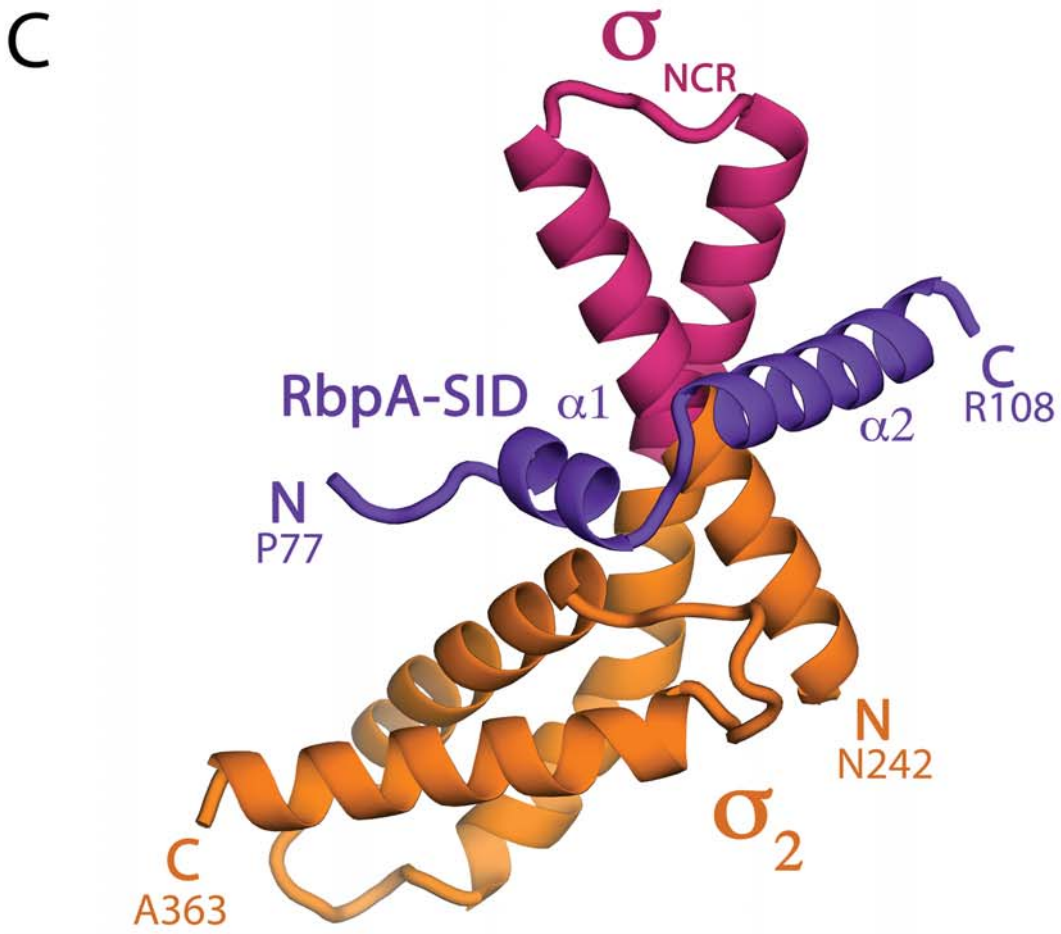
Fig. 3. Structural model of *Mtb* RbpA-SID on an RPo complex. (A) A model of the RbpA-SID/RPo complex was generated by superimposing conserved regions of *Mtb* σ^{A}_{2} (from the *Mtb* $RbpA_{SID}/\sigma^{A}_{2}$ structure) with *Taq* σ^{A} in an RPo model. Protein and DNA elements are colored as indicated. RNAP is shown as a molecular surface; RbpA-SID/σ^{A}_{2} is shown in ribbon; and DNA is shown as a phosphate backbone worm. (B) A magnified view of the RbpA/RPo model shows that RbpA is positioned to make contacts with the DNA upstream of the -10 promoter element. RbpA amino acids R79, located in the RCD-SID linker, and M84, located in α1, are in position to contact the nt-strand phosphate backbone at the -13/-14 position. RbpA-M84 makes non-polar contacts with *Mtb* σ^{A} residue K334, which interacts with -12T at the beginning of the transcription bubble. RbpA sequence reveals three conserved, positively charged lysine residues K76, K74, K73 (represented by purple dots) located in the RCD-SID linker that would be well positioned to interact with the negatively charged DNA phosphate backbone.

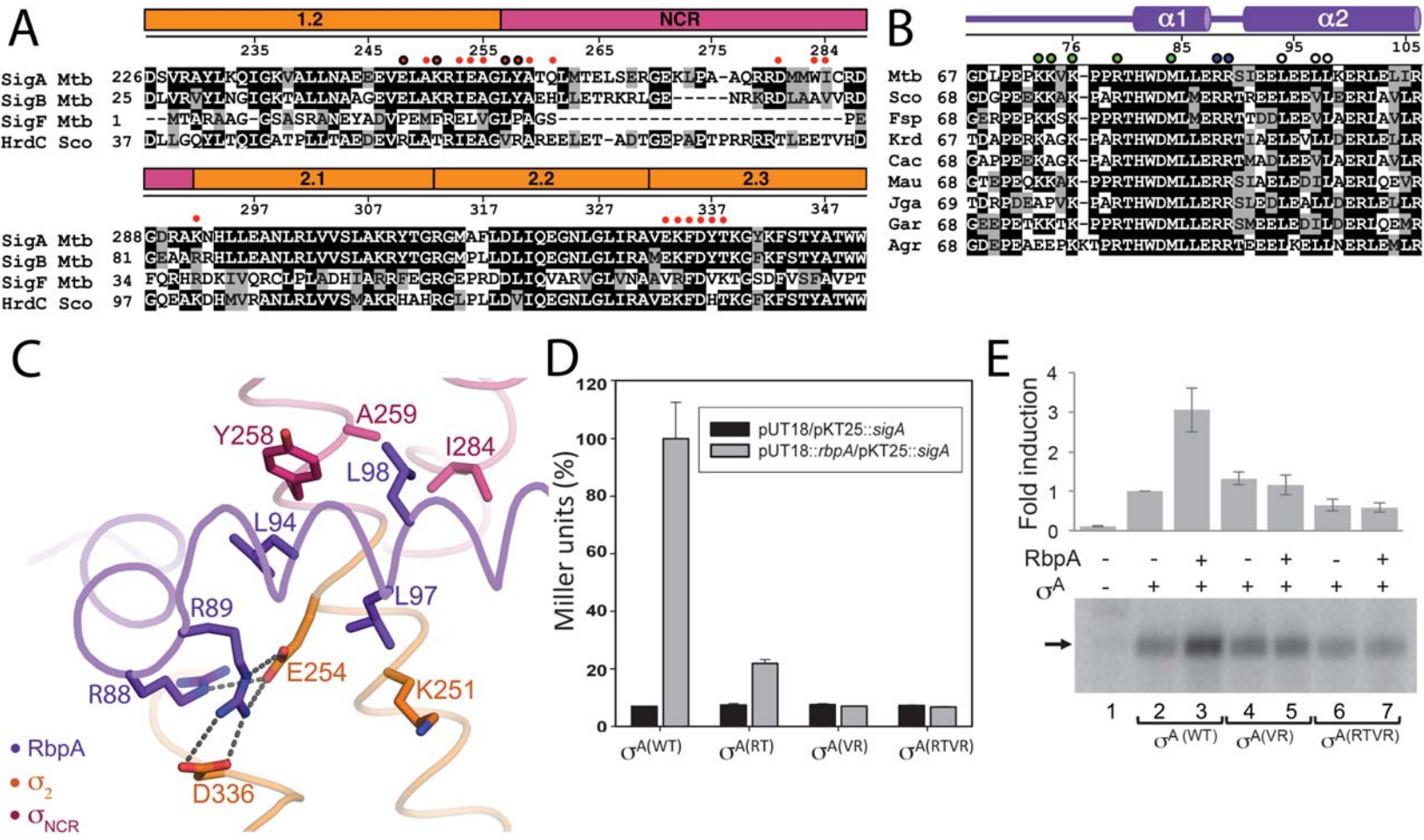
Fig. 4. RbpA directly contacts the DNA and increases affinity of holenzyme to promoter DNA. (A) RbpA cross-links to fork junction promoter DNA. Reactions contained VFJ-28 DNA 5'-labeled on the non-template strand, core *Mbo* RNAP (E) at 200 nM, σ^{A} (1 μM), and native or SUMO-fused derivatives of RbpA (2 μM), as indicated. Complexes were allowed to form at 37°C for 15 min before treatment with formaldehyde. Cross-linked species were separated by 4-12% SDS-PAGE and visualized by phosphorimaging. The *vapB10* promoter-based fork junction DNA template VFJ-28 is illustrated, indicating the -10 element (blue) and labeled non-template strand (asterisk). (B) RbpA increases

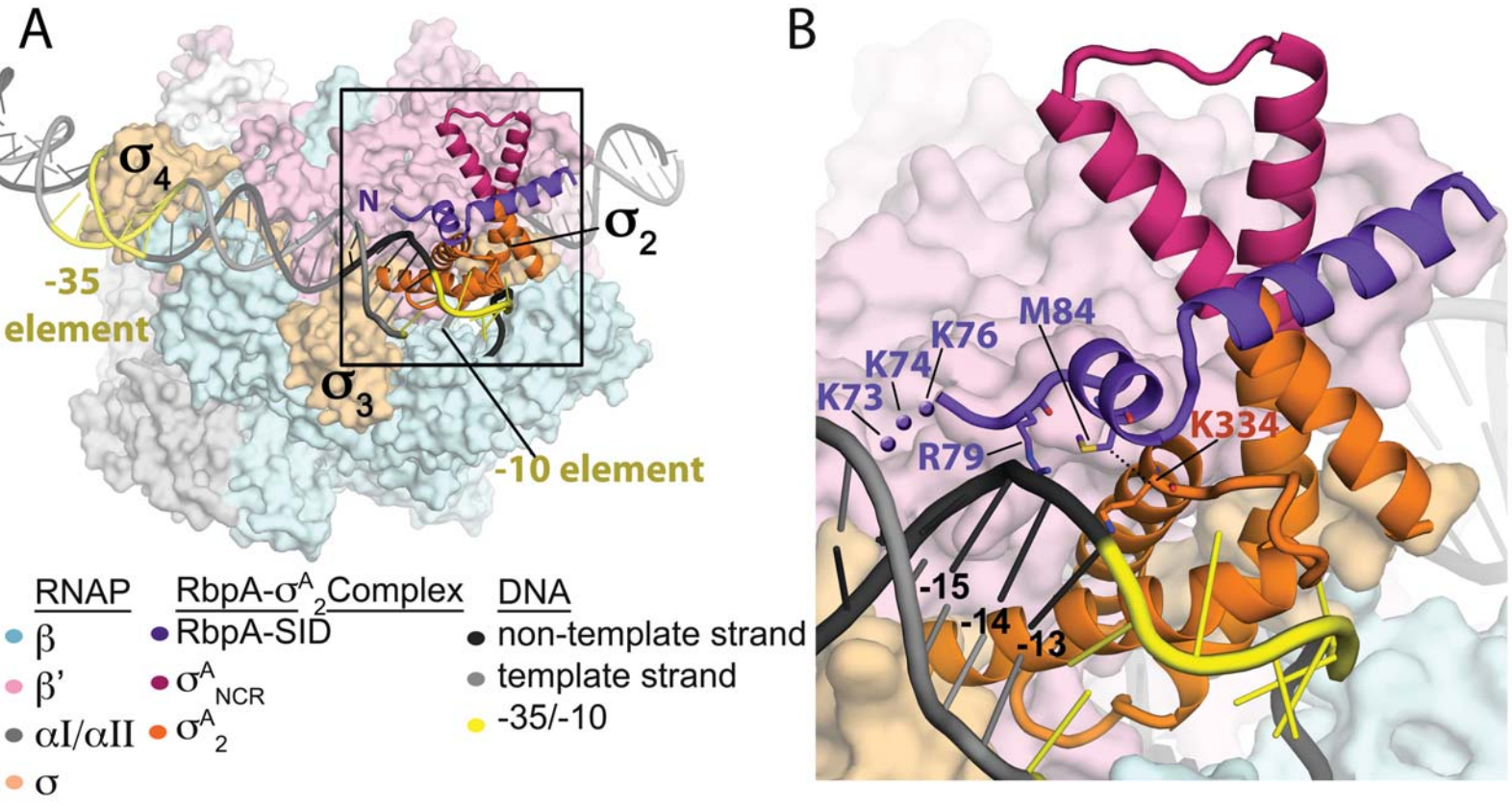
RNAP affinity for promoter DNA. Representative fluorescence anisotropy binding curves show RNAP binding to Cy3-labeled *vapB10* promoter DNA (Fig. S4A). Standard error bars and average K_Ds are based on triplicate trials. Inset shows average fold activation (normalized to no RbpA). WT RbpA increases affinity of RNAP for promoter DNA while RbpA-R79A has little-to-no effect on binding. Error bars are based on average fold change from nine experiments. (C) RbpA residues R79 and M85, predicted to bind DNA, are important for *in vitro* transcription activation. Multiple round *in vitro* run-off transcription reactions containing *Mbo* RNAP (50 nM), σ^A (250 nM), and RbpA (500 nM, or 1.25 μM) using the *vapB10* promoter as template. A graphical representation based on triplicate data sets, with standard deviation indicated, is illustrated above, normalized to the data obtained with σ^A in the absence of RbpA.

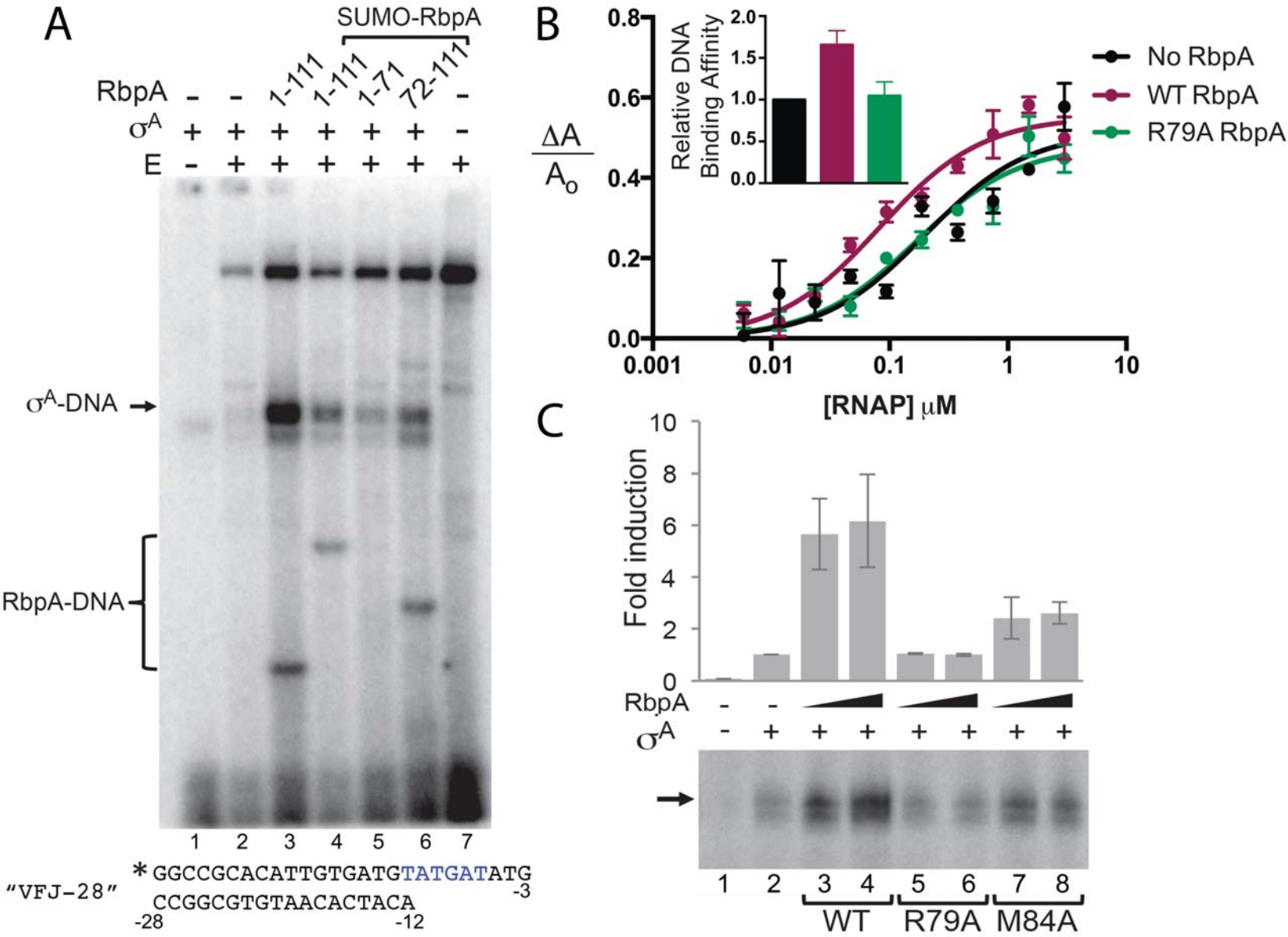

A
Mtb σA
224
σA2
369
σA3
489
σA4
1
1.1
1.2
NCR
2.1
2.2
2.3
2.4
3.0
3.1
3.2
4.1
4.2
528
242
363
Mtb RbpA
26
66
81
1
RCD
BL
SID
111
77
108


B
Fold induction
5
4
3
2
1
0
RbpA
σA
1-111
1-111
1-71
72-111
1 2 3 4 5 6 7 8 9 10 11
RbpA
SUMO-RbpA


C
σ NCR
RbpA-SID
α1
α2
C
R108
N
P77
N
N242
σ2
C
A363


A
1.2
NCR
2.1
2.2
2.3
SigA Mtb
SigB Mtb
SigF Mtb
HrdC Sco
B
α1
α2
Mtb
Sco
Fsp
Krd
Cac
Mau
Jga
Gar
Agr
C
A259
Y258
I284
L98
L94
L97
R89
R88
E254
K251
D336
RbpA
σ2
σNCR
D
Miller units (%)
pUT18/pKT25::sigA
pUT18::rbpA/pKT25::sigA
σA(WT)
σA(RT)
σA(VR)
σA(RTVR)
E
Fold induction
RbpA
σA
σA (WT)
σA(VR)
σA(RTVR)


A
σ4
N
σ2
-35
element
σ3
-10 element
RNAP
β
β'
αI/αII
σ
RbpA-σA2Complex
RbpA-SID
σA NCR
σA2
DNA
non-template strand
template strand
-35/-10
B
M84
K76
K74
K73
R79
K334
-15
-14
-13


A
SUMO-RbpA
RbpA
1-111
1-111
1-71
72-111
σA
E
σA-DNA
RbpA-DNA
1 2 3 4 5 6 7
"VFJ-28"
* GGCCGCACATTGTGATGTATGATATG
CCGGCGTGTAACACTACA
-28
-12
-3
B
ΔA / Ao
Relative DNA Binding Affinity
No RbpA
WT RbpA
R79A RbpA
[RNAP] μM
C
Fold induction
RbpA
σA
1 2 3 4 5 6 7 8
WT
R79A
M84A

SI Materials and Methods:

Protein Expression and Purification. For crystallography, σ^{A}_{2} (codons 224-364) and *rbpA* (codons 1-111) from *Mtb* were chemically synthesized as a single DNA fragment with *Eco* codon usage and independent T7 promoter and translation initiation signals for *rbpA* (Genscript), cloned as a *Bam*HI-*Hin*dIII fragment into a His_6 pET SUMO expression vector, and transformed into *Eco* BL21(DE3) cells (Novagen). Transformed cells were grown at 37 °C in the presence of 50 μg/ml kanamycin to an OD_{600} of 0.6, at which point the temperature was lowered to 30 °C, and protein expression was induced with 500 μM IPTG for 3 h. Cells were harvested by centrifugation and resuspended in Lysis Buffer [20 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole, 5% glycerol (v/v), 0.5 mM β-mercaptoethanol] supplemented with 1 mM PMSF and protease inhibitor cocktail. Cells were lysed by French press (Avestin), and lysate was cleared by centrifugation. Clarified lysate was loaded on a Hi-Trap IMAC Ni^{2+}-chelating column (GE Healthcare) and eluted with Lysis Buffer containing 250 mM imidazole. The elution was directly loaded on a size exclusion column (SuperDex-200 16/16, GE healthcare) equilibrated with 20 mM Tris-HCl, pH 8, 0.5 M NaCl, 5% glycerol (v/v), 1 mM DTT, 1 mM EDTA. The sample was concentrated to 15 mg/ml by centrifugal filtration and stored at -80 °C. For fluorescence anisotropy, *Mtb* RbpA and *Mtb* RbpA-R79A were cloned into a His_6 pET28-based vector, and the plasmid was transformed into *Eco* Rosetta 2 BL21(DE3) cells. The proteins were overexpressed and purified using the same buffers and techniques as the RbpA/σ_2 co-expressed complex. Core RNAP, σ^{A}, and CarD were over-expressed and purified using previously described method (14).

For *in vitro* transcription and cross-linking assays, pSX500, which contains *Mtb rbpA* cloned in pET20b, as well as mutant derivatives, were overexpressed in *Eco* BL21(DE3) pLysS following a 15 min cold-shock of cultures (OD_{600} 0.4-0.7) on ice, which improved protein solubility, and induction with 1 mM IPTG for 3h. Protein was purified by ion exchange chromatography (HiTrap Q Sepharose™ Fast Flow and Mono Q 5/50 GL, GE Healthcare) and size exclusion chromatography (HiLoad 16/60 Superdex 200, GE Healthcare). To generate N-terminal SUMO fusions of *Mtb* RbpA (amino acid residues 1-111) and truncated derivatives (amino acid residue 1-71 and 72-111), RbpA derivatives were cloned into pET SUMO (Invitrogen) as *Bam*HI–*Hin*dIII fragments, overexpressed in *Eco* BL21(pLysS) as described above, and purified by a sequential Ni-affinity chromatography (Ni^{2+}-charged iminodiacetic acid sepharose; Sigma-Aldrich), ion exchange chromatography (Mono Q 5/50 GL, GE Healthcare) and size exclusion chromatography (HiLoad 16/60 Superdex 200, GE Healthcare). σ^A and mutant derivatives ($\sigma^{A(VR)}$ and $\sigma^{A(RTVR)}$) were cloned into pET15b and purified from *Eco* BL21(pLysS) following induction with 1 mM IPTG for 3 h at 30°C. Cell pellets were resuspended in ice-cold Lysis Buffer and following cell lysis by sonication, clearing of cell lysates by centrifugation and binding to a Ni column (see above), protein was eluted by thrombin-cleavage (Sigma-Aldrich), then further purified by size-exclusion chromatography as outlined above.

Crystallization of RbpA/σ^A_2 Complex. Crystals of the RbpA/σ^A_2 complex were grown at 22 °C via sitting drop vapor diffusion against a reservoir solution of 0.1 M Tris, pH 8.5, and 0.5 M ammonium sulfate, at a protein concentration of 15 mg/ml. Drops were set up with a 1:1 ratio of His_6-SUMO-σ^A_2-RbpA:crystallant. Crystals measuring approximately 50 μm took

3 days to grow and were cryoprotected in reservoir solution plus 30% (v/v) ethylene glycol.

MALDI-TOF-MS of Crystals. One to two crystals were removed from the mother liquor and briefly washed in cold water and subsequently dissolved in matrix solution, which consisted of a saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA) in a 1:3:2 (v/v/v) mixture of formic acid/water/isopropanol (FWI). An aliquot of 0.5 µl of the protein-matrix solution was transferred onto a MALDI plate pre-coated with an ultrathin layer. The ultrathin layer was prepared and laid on the plate as previously published (30, 31). The sample spots were washed for a few seconds with 2 µl of cold 0.1% aqueous trifluoroacetic acid (TFA) solution. MALDI spectra were acquired in linear, delayed extraction mode using a Spiral TOF JMS-S3000 (JEOL, Tokyo, Japan) equipped with a Nd:YLF laser, delivering 10-Hz pulses at 349 nm. Delayed extraction time was set at 1 ms, and acquisition was performed with a sampling rate of 2 ns. Each MALDI spectrum corresponded to an average of 500 scans. Mass calibration was performed using a technique of pseudo-internal calibration wherein a few shots on a nearby calibrant spot were collected and averaged with the sample shots into a single spectrum. The spectra were processed and analyzed using MoverZ (Proteometrics, LLC). MALDI-TOF analysis of the crystals revealed that both $\sigma^{A}{}_{2}$ and RbpA were proteolytically degraded at their N-termini. The crystals were composed of mixtures of σ^{A} missing up to 18 N-terminal residues, and two sets of RbpA fragments. Fragments of σ^{A} detected included: S224-R364, A230-Q362, Y231-A363, L233-R364, L240-Q362, N242-R364. Two groups of N-terminal variants of RbpA were detected: One group contained most of the N-terminal region,

including all of the RCD (beginning at M1, R10, S15, E17, R20). The second group lacked the RCD and included N-terminal variants starting at R57, G59, L64, G67, K74). All RbpA variants contained the native C-terminal residue G111.

Data Collection and Refinement of σ^{A}_{2}-RbpA Complex. X-ray diffraction data were collected at the Brookhaven National Synchroton Light Source X29 beamline and integrated and scaled using HKL2000 (32). Electron density maps were generated using molecular replacement with an *Mtb* σ^{A}_{2} homology model (ExPASy SWISS MODEL) based on a previously solved structure of the corresponding domain from *Taq* σ^{A} [1KU2; (4)]. The model was built using reiterative cycles of manual building with COOT (33) and refinement with Phenix (34) (Table S2). The final model included residues 242-363 of σ^{A} and 77-108 of RbpA. The RCD of RbpA, which lacked density, was presumed disordered. The PDBePISA server (http://www.ebi.ac.uk/pdbe/pisa/) was used to calculate intermolecular buried surface areas (35).

Fluorescence Anisotropy. Fluorescence measurements were performed using an Infinite M1000pro plate reader (Tecan) in a 384 well plate with a final reaction volume of 25 uL. Holoenzyme was formed at 37°C using a 1:1.5 ratio of core RNAP:σ^{A}. All proteins were dialyzed and then serially diluted in Anisotropy Buffer consisting of 100 mM potassium-glutamate, 10 mM HEPES, pH 8, 10 mM $MgCl_2$, 0.1 mM DTT, 0.02% TWEEN 20, and 5% (v/v) glycerol. When used, RbpA was added to RNAP in 5-fold excess. Protein complexes were serially diluted in Anisotropy Buffer to obtain RNAP concentrations ranging from 5

nM to 3 μM. For assays with Cy3-labled single stranded DNA extending from the -12 to +1 (Tri-link, Table S3), DNA was diluted in Anisotropy Buffer and added to the protein mixture at a final concentration 200 nM. For assays performed with double stranded *vapB10* DNA extending from the -36 to +10, Cy3-labeled non-template strand DNA (Tri-link) was annealed to unlabeled template strand DNA (Oligos Etc.), diluted in Anisotropy Buffer, and added to the protein mixture at a final concentration of 10 nM. Measurements of RNAP binding to double-stranded DNA was noisy without the addition of CarD [a result of the unstable open complex and short half-life of *M. bovis* RNAP; (14)], so these assays included CarD (in 3-fold excess over RNAP). Data were analyzed using Prism software. Assays with single-stranded DNA were performed in triplicate. For assays with duplex DNA, three experiments composed of side-by-side triplicate trials for each condition were performed.

Formaldehyde cross-linking. Fork-junction DNA was prepared by labeling the 5' end of the non-template oligonucleotide using [γ^{32}P]ATP, annealing with 3-fold molar excess of template strand oligonucleotide, followed by purification using a NAP-5 column (GE Healthcare). The fork-junction template VFJ-28 contained an 18 bp double stranded region from -28 to -12 followed by a single-stranded extension from -11 to -3 on the non-template strand (Fig. S3A). In control experiments, an anti-consensus -10 sequence was used (VFJ-28anti), changing the -10 element from TATGAT to AGTGAC. σ^A was mixed with *Mtb* RbpA and mutant and truncated derivatives on ice for 10 min before the addition of core *M. bovis* (*Mbo*) RNAP. Reactions were then set up containing 400 nM σ^A, 200 nM core RNAP, various concentrations of RbpA, ~10 nM fork junction DNA in binding buffer [40 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 0.6 mM EDTA, 20% (v/v) glycerol, 0.8 mM potassium

phosphate (pH 7.5), 0.75 mM DTT, and 0.125 mg/ml BSA] and equilibrated at 37□C for 15 min. Cross-linking was initiated by the addition of formaldehyde (60 mM final) and stopped by the addition of 2x stop buffer (4% SDS, 20% glycerol, 10% β-mercaptoethanol, 0.004% bromophenol blue, 125 mM Tris-HCl, pH 6.8) after 2 min. Samples were loaded on Precast NuPAGE® 4-12% Bis-Tris gels (Invitrogen) at 100V for 3 h at 10□C, dried, then analyzed by phosphorimagery.

***In vitro* transcription.** Multi-round run-off and abortive initiation *in vitro* transcription assays were carried out using core *Mbo* RNAP, purified *Mtb* σ^A and *Mtb* RbpA or mutant and truncated derivatives essentially as described (15). The *vapB10* promoter template was amplified by PCR using the oligonucleotides vapB10_F (5'-GCGCTGAAGAGGGCGTTGCAC) and vapB10_R (5'-TTCAGCAGGAGGCGGATCAG). Multi-round assays, giving rise to a 109 nt run-off transcription product, were performed at 37°C and contained 5 nM template in reaction buffer [40 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 0.6 mM EDTA, 20% (v/v) glycerol, 0.8 mM potassium phosphate (pH 7.5), 0.75 mM dithiothreitol, and 0.125 mg/ml BSA]. Reactions were pre-incubated at 37°C for 10 min before the addition of NTP mix [200 □M ATP, GTP, CTP, 50 μM UTP, including 5 μCi [α-^{32}P] UTP (>800 Ci/mmol), GE Healthcare], and incubated for 10 min. Reactions were halted by addition of an equal volume of *in vitro* loading dye [80% (w/v) formamide, 0.01 % (w/v) xylene cyanol, 0.01 % (w/v) bromophenol blue], separated on denaturing 8% polyacrylamide gels, and quantified by phosphorimaging. Abortive initiation assays were performed at 37 °C in a buffer containing 10 mM Tris-HCl, pH 8.0, 100 mM potassium-glutamate, 10 mM $MgCl_2$, 0.1 mM EDTA, 0.1

mM DTT, and 50 μg/mL BSA. Core RNAP (50 nM), σ^A (250 nM) and RbpA (500 nM, when used) were preincubated at 37°C for 5 min. DNA (10 nM) was added and the reaction was incubated for 15 min to allow RP_o to form. Abortive transcription was initiated with an ApU dinucleotide primer (250 μM), [α-^{32}P]GTP (1.25 μCi), and unlabeled GTP (50 μM). After 10 min, reactions were quenched with 2x stop buffer (8M urea, 0.5x TBE, 0.05% bromophenol blue, 0.05% xylene cyanol) and separated on a 23% urea-polyacrylamide gel. Abortive products were visualized by phosphorimagery and quantified using Image J.

Supplementary Figure Legends

Fig. S1. Molecular interactions in the *Mtb* RbpA/ σ^{A_2} complex

(A) Stereoview of selected portion of a simulated annealing composite omit electron density map (2mFo-DFc), calculated with RbpA omitted, contoured to 1.0 σ. Map was calculated in Phenix (34) with a starting temperature of 5000 K. Clear electron density for RbpA is shown. RbpA (purple) and σ^{A_2} (orange with the NCR colored cranberry) are shown in stick, highlighting the hydrophobic interactions between RbpA residues L94 and L98 with $\sigma^{A}{}_{NCR}$-Y258. (B) Schematic highlights polar and ionic interactions between RbpA-SID and σ^{A}. Interactions between residues are indicated by lines, with ionic interactions colored in red, hydrogen bonds colored in gray, and hydrogen bonds mediated by waters colored in blue. C. Schematic shows non-polar (van der Waals) interactions between RbpA-SID and σ^{A}.

Fig. S2. BTH analysis of *Sco* RbpA wild-type (WT) and RbpA-V98A/L99A interactions with *Sco* σ^{HrdB}. (A) β-galactosidase assays were conducted with *E. coli* BTH101 strains containing: C1, pUT18 and pKT25; C2, pUT18 and pKT25-*hrdB*; C3, pUT18-*rbpA* and pKT25; RbpA (WT), pUT18-*rbpA(Sco)* and pKT25-*hrdB*; RbpA (V98A/L99A), pUT18-*rbpA(Sco)*-V98A/L99A and pKT25-*hrdB. rbpA* coding sequences were fused to the T18 subunit of *Bordetella pertussis* adenylate cyclase in pUT18, whereas the sigma factor genes (domains σ_2, σ_3 and σ_4) were fused to the T25 subunit in pKT25 (15). β-galactosidase assays were performed in triplicate (standard deviations indicated) and results are presented as % Miller units relative to results obtained with wild-type *Sco* RbpA). (B) Western analysis of total cell extracts of *E. coli* BTH101 strains containing: RbpA (WT), pUT18-*rbpA(Sco)* and pKT25-*hrdB*; RbpA (V98A/L99A), pUT18-*rbpA(Sco)*-V98A/L99A and pKT25-*hrdB*. Extracts were prepared at the time of β-galactosidase assay by precipitation with 10% (v/v) trichloroacetic acid, and T18-RbpA fusion proteins were detected using an anti-adenylate cyclase toxin antibody (3D1, Santa Cruz Biotechnology).

Fig. S3. Formaldehyde cross-linking indicates specificity for promoter DNA interactions and a requirement for holoenzyme to form RbpA-DNA cross-links.

(A) The *vapB10* promoter is indicated with the -10 element highlighted in blue. VFJ-28 is the fork junction template based on the *vapB10* promoter. VFJ-28anti is identical to VFJ-28 apart from changes at three positions in the -10 element (underlined). (B) Formaldehyde cross-linking reactions contained VFL-28 or VFK-28anti, core *Mbo* RNAP (E, 200 nM), σ^A (400 nM), and native or SUMO-fused derivatives of RbpA (at increasing molar concentrations relative to E), as indicated. Cross-linked species were separated by 4-12% SDS-PAGE and visualized by phosphorimaging.

Fig. S4. Fluorescence anisotropy assays

(A) The Cy3-labeled *vapB10* double-stranded DNA probe (-36 to +1) used for anisotropy experiment with the -10 element highlighted in blue. (B) Multi-round abortive initiation transcription assays on the *vapB10* promoter showing activation by RbpA. Standard error bars based on 4 trials.

Fig. S5. RbpA R80 and M85 are required for normal growth of *S. coelicolor.*

S. coelicolor S129 (Δ*rbpA*), containing the integrative vector pSETΩ, and S129 derivatives containing wild-type (*rbpA*-WT) or mutant derivatives of *rbpA* (*rbpA*-R80A or *rbpA*-M85A) cloned in pSETΩ, were grown on mannitol-soya agar containing spectinomycin (25μg/ml). The agar plate was photographed after incubation at 30°C for 3 days.

Fig. S6. Proposed RbpA binding sites on RNAP core are incompatible with current structure.

Protein and DNA elements are colored as in Fig. 2. (A) The proposed RbpA interacting residue R381 (red) on RNAP β subunit was predicted through crosslinking to bind K74 in the BL (27). However, RbpA-P77, just 4 residues away from K74, is 53 Å away from βR381 (direct Cα-Cα distance). (B) The RbpA-RCD binding regions predicted by iron-babe cleavage (green) (18) are located at least 80Å away from the beginning of the RCD-SID linker (measured as a straight line from the Cα of P77). The clamp region, proposed in this study to be the region RbpA would more likely interact with, is highlighted by arrow.

Table S1. Bacterial strains, plasmids and PCR primers used in this study

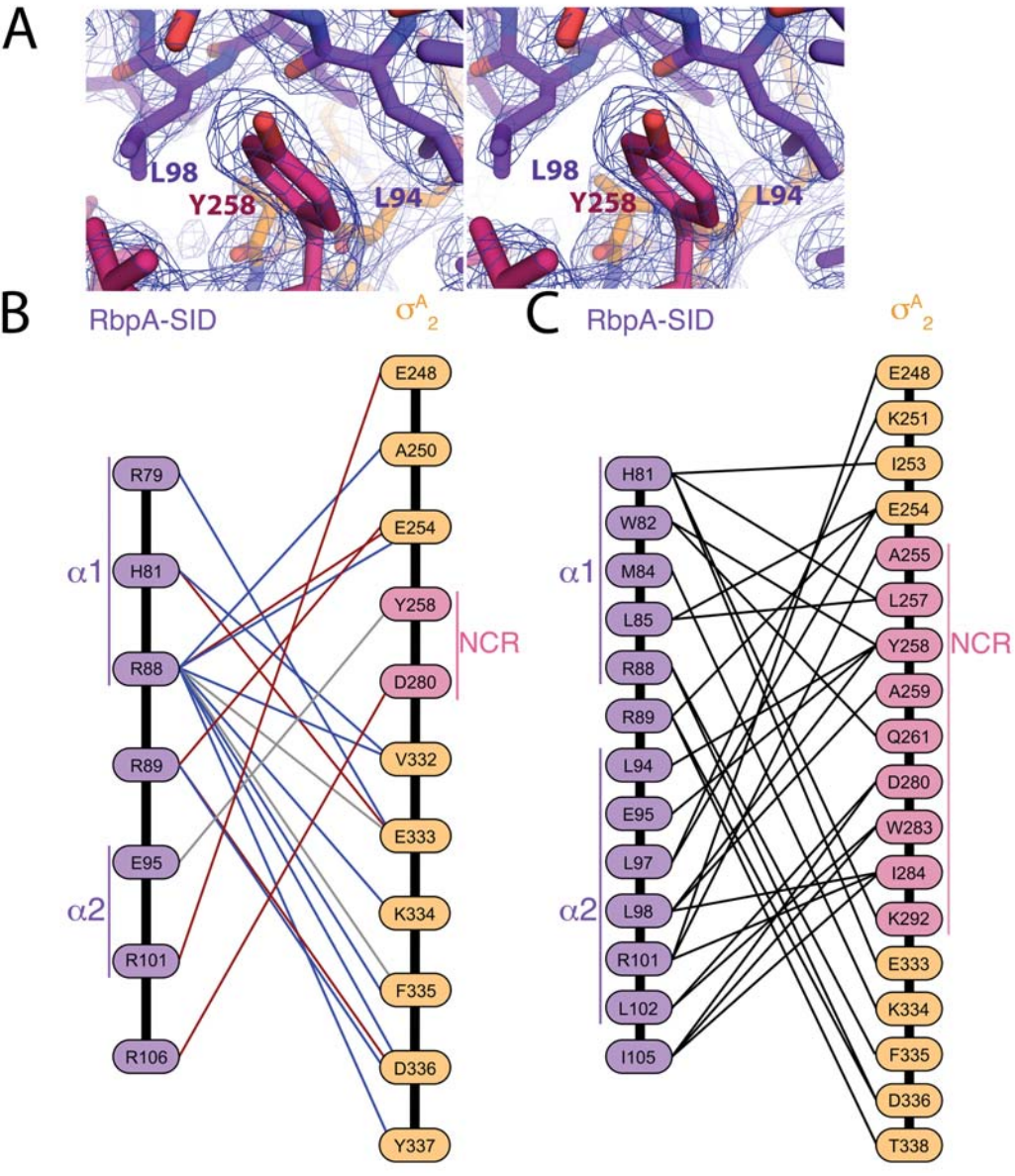
Strain or plasmid	Relevant genotype/comments	Source/reference
Strains		
***S. coelicolor* A3(2)**		
S129	J1981 Δ*rbpA::aprhyg* (AprHyg^R)	(15)
E. coli		
ET12567 (pUZ8002)	*dam*, *dcm, hsdM.* pUZ8002 is a non-transmissible derivative of RK2 (Cm^R, Km^R)	(36)
BL21 (DE3)	*E. coli* B F$^-$ *ompT hsdS*($r_B^-m_B^-$) *dcm gal* λ(DE3) *endA* Hte (Tet^R)	(37)
BL21 (DE3) pLysS	BL21(DE3) pLysS (Cm^R)	(37)
BTH101	*cya-99* (Spc^R) used for bacterial two-hybrid analysis	(38)
Rosetta 2	BL21(DE3) expressing rare codon tRNA genes	Novagen
Plasmids		
pBluescript II SK$^+$	*E. coli* cloning vector; *ori* pUC18 (Amp^R)	
pKT25	Two-hybrid vector; T25 fragment of *B. pertussis* CyaA for N-terminal fusions (Km^R)	(38)
pUT18	Two-hybrid vector; T18 fragment of *B. pertussis* CyaA for C-terminal fusions (Amp^R)	(38)
pKT25-*hrdB*	*hrdB* (codons 211-511) fused to T25 in pKT25 (Km^R)	(15)
pKT25-*sigA*	*Mtb sigA* (codons 1-528) fused to T25 in pKT25 (Km^R)	(15)
pKT25-*sigA*(RT)	*sigA* (E248R/ K251T, codons 1-528) fused to T25 in pKT25 (Km^R)	This study
pKT25-*sigA*(VR)	*sigA* (L257V /Y258R , codons 1-528) fused to T25 in pKT25 (Km^R	This study
pKT25-*sigA*(RTVR)	*sigA* (E248R/K251T/L257V/Y258R, codons 1-528) fused to T25 in pKT25 (Km^R	This study
pUT18-*rbpA(Mtb)*	*Mtb rbpA* (codons 1-111) fused to T18 in pUT18 (Amp^R)	(15)
pUT18-*rbpA(Sco)*	*Mtb rbpA* (codons 1-124) fused to T18 in pUT18 (Amp^R)	(15)
pSETΩ	Integrative cloning vector (Spc^R)	(15)
pSX530	pSETΩ containing at the *Bam*HI site a *Bgl*II-fragment including *rbpA.*	(15)
pSX530(R80A)	As pSX530 but with *rbpA-R80A* mutation	This study
pSX530(M85A)	As pSX530 but with *rbpA-M85A* mutation	This study
pET15b	*E. coli* expression vector (His$_6$-tagged; Amp^R)	Novagen
pET15b-*sigA*	*Mtb sigA* cloned in pET15b as a *Nde*I-*Bgl*II fragment	This study
pET15b-*sigA*(RT)	*Mtb sigA* (E248R/ K251T) cloned in pET15b as a *Nde*I-*Bgl*II fragment	This study
pET15b-*sigA*(VR)	*Mtb sigA* (L257V /Y258R) cloned in pET15b as a *Nde*I-*Bgl*II fragment	This study
pET15b-*sigA*(RTVR)	*Mtb sigA* (E248R/K251T/L257V/Y258R) cloned in pET15b as a *Nde*I-*Bgl*II fragment	This study
pET20b	*E. coli* expression vector (native; Amp^R)	Novagen
pSX500	pET20b containing *M. tuberculosis rbpA cloned as Nde*I-*Bam*HI *fragment*	(15)
pSX500(R79A)	pET20b containing *M. tuberculosis rbpA*-R79A cloned as *Nde*I-*Bam*HI fragment	This study
pSX500(M84A)	pET20b containing *M. tuberculosis rbpA*-R79A cloned as *Nde*I-*Bam*HI fragment	This study
pET SUMO	*E. coli* expression vector (N-terminal His$_6$/SUMO fusion; Km^R)	Life Technologies
pET28	*E. coli* expression vector (His$_6$-tagged; Km^R)	Novagen
pET28-*rbpA*	*Mtb rbpA* cloned into pET28.	This study
pET28-*rbpA*(R79A)	As pET28-*rbpA* but with R79A allele	This study
pSUMO-*sigA*/*rbpA*	*Mtb sigA* (codons 224-364) and *rbpA* (codons 1-111) chemically synthesized and cloned as a *Bam*HI-*Hin*dIII fragment into a pET SUMO, generating a SUMO-*sigA* fusion	This study
pSUMO-*rbpA*(1-111)	pET SUMO containing *Mtb rbpA* (codons 1-111) fused to C-terminus of SUMO as a *Bam*HI-*Hin*dIII fragment	This study
pSUMO-*rbpA*(1-71)	pET SUMO containing *Mtb rbpA* (codons 1-71) fused to C-terminus of SUMO as a *Bam*HI-*Hin*dIII fragment	This study
pSUMO-*rbpA*(72-111)	pET SUMO containing *Mtb rbpA* (codons 72-111) fused to C-terminus of SUMO as a *Bam*HI-*Hin*dIII fragment	This study
PCR primers	**Primer sequence (5' to 3') (restriction sites indicated in bold)**	
Plasmid constructs		
pSUMO-RbpA(1-111)	**GGATCCCATATG**GCTGATCGTGTCCTGAGG **AAGCTT**TCAGCCG CGCCGACGTGACCGAATG	
pSUMO-RbpA(1-71)	**GGATCCCATATG**GCTGATCGTGTCCTGAGG **AAGCTT**TCACTCGGGCAGGTCGCCCTCGATCAG	
pSUMO-RbpA(72-111)	**GGATCCCATATG**GAGCCGAAGAAGGTTAAGCCG	

AAGCTTTCAGCCG CGCCGACGTGACCGAATG

Table S3: Average K_Ds of RNAP binding to Cy3 labeled ssDNA (-12 to +1) with and without RbpA

Promoter	Sequence	No RbpA K_D (μM)	RbpA K_D (μM)
Consensus	*TATAATGGGAAGG	0.38 ± 0.07	0.33 ± 0.04
AP3	*TAGACTGGCAGGG	2.24 ± 0.41	1.88 ± 0.31
vapB	*TATGATATGGTGT	1.62 ± 0.26	1.77 ± 0.15

Average K_Ds of RNAP binding to Cy3 labeled ssDNA (-12 to +1) were determined using fluorescence anisotropy. The addition of RbpA to the assay has no significant effect on RNAP binding to ssDNA composed of the -10 and discriminator elements.
-10 element is highlighted in blue. Error is based off of triplicate trials.


A
L98
Y258
L94
L98
Y258
L94
B
RbpA-SID
σA2
R79
H81
R88
R89
E95
R101
R106
α1
α2
E248
A250
E254
Y258
D280
V332
E333
K334
F335
D336
Y337
NCR
C
RbpA-SID
σA2
H81
W82
M84
L85
R88
R89
L94
E95
L97
L98
R101
L102
I105
α1
α2
E248
K251
I253
E254
A255
L257
Y258
A259
Q261
D280
W283
I284
K292
E333
K334
F335
D336
T338
NCR

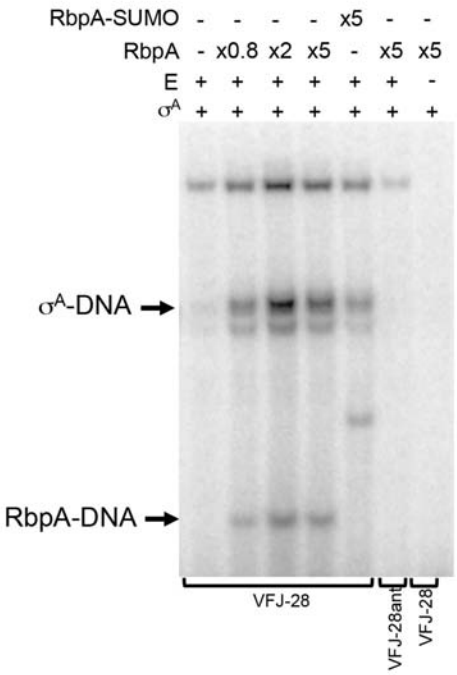
A

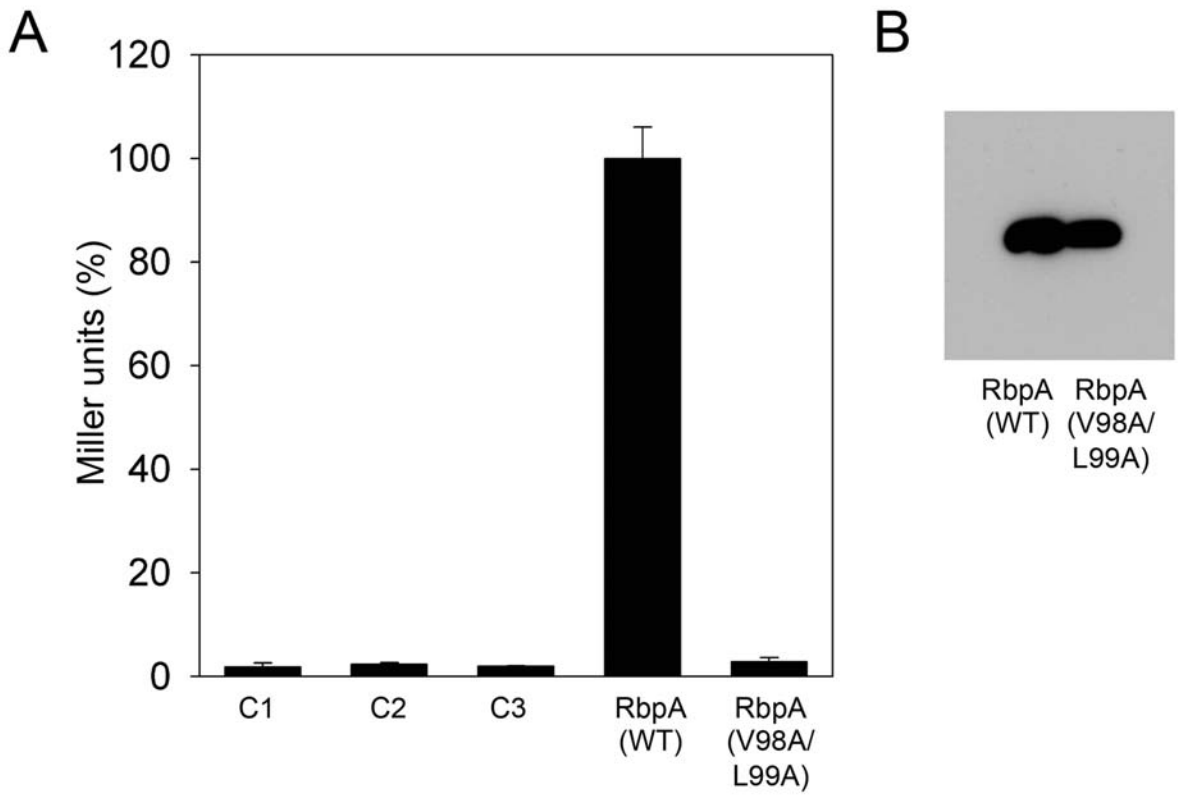
vapB10p 5′ GGCCGCCCGGGCCGCACATTGTGATGTATGATATGGT*G* 3′
3′ CCGGCGGGCCCGGCGTGTAACACTACATACTATACCAC 5′

“VFJ-28” 5′GGCCGCACATTGTGATGTATGATATG 3′
3′CCGGCGTGTAACACTACA 5′

“VFJ-28anti” 5′GGCCGCACATTGTGATGAGTGACATG 3′
3′CCGGCGTGTAACACTACT 5′

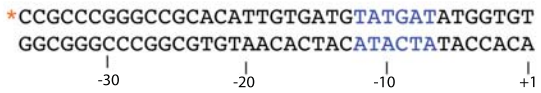
B



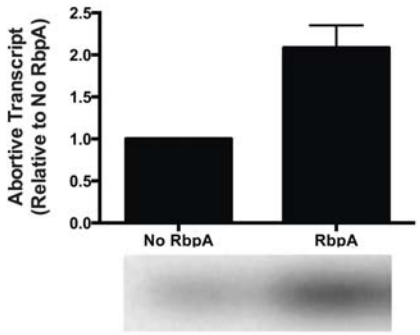

A
120
100
80
60
40
20
0
Miller units (%)
C1
C2
C3
RbpA (WT)
RbpA (V98A/ L99A)
B
RbpA (WT)
RbpA (V98A/ L99A)

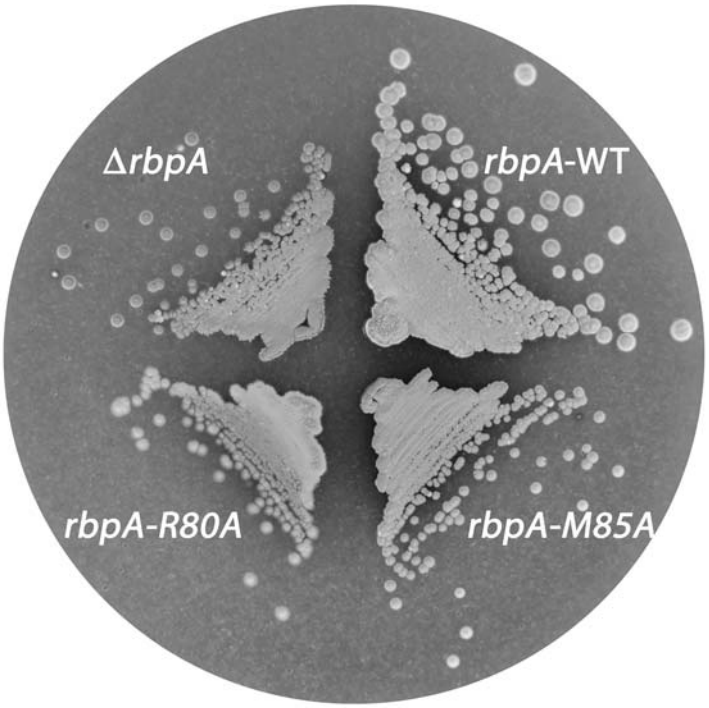
A

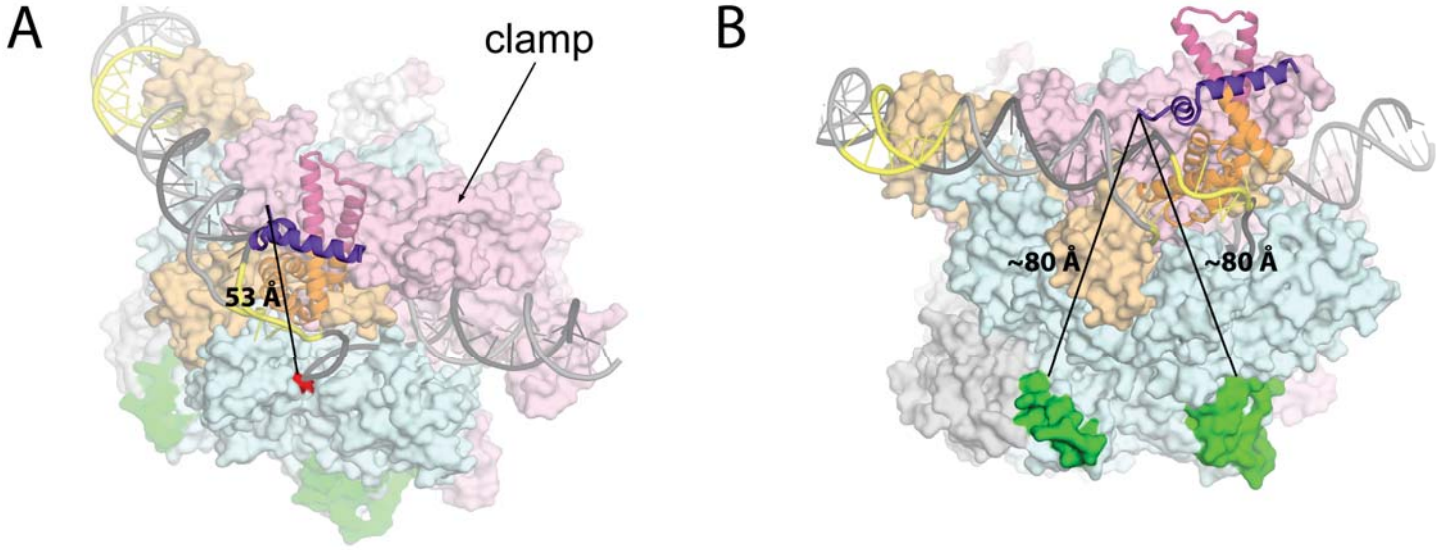
vapB10 dsDNA



B




ΔrbpA
rbpA-WT
rbpA-R80A
rbpA-M85A


A
clamp
53 Å
B
~80 Å
~80 Å